

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

Memo No. 326

December 1974

THE RECOGNITION OF SHARP, CLOSELY SPACED EDGES

by

David Marr

ABSTRACT

The recognition of sharp edges from edge- and bar-mask convolutions with an image is studied for the special case where the separation of the edges is of the order of the masks' pixel-widths. De-smearing techniques are employed to separate the items in the image. Attention is also given to parsing de-smear mask convolutions into edges and bars; to detecting edge and bar terminations; and to the detection of small blobs.

Work reported herein was conducted at the Artificial Intelligence Laboratory, a Massachusetts Institute of Technology research program supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under Contract number N00014-70-A-0362-0005.

Introduction

The thesis was advanced elsewhere (Marr 1974a) that the purpose of low-level vision should be to compute a very low-level symbolic description of an image, using an appropriately powerful set of symbols and methods; and that subsequent processes should have access only to that description. An explicit low-level vocabulary was defined, and methods were given by which a low-level symbolic description may be computed from an image (Marr 1974b). These methods operated on the assumption that the separation of the edges in the image was large compared with the size of the masks that were used to measure the first (edge-shaped masks) and second (bar-shaped masks) directional derivatives of the intensity in the image. In this article, that assumption is relaxed, and the recognition of very closely-spaced edges is considered. The methods that are necessary in this situation are somewhat unsatisfactory, and the problem is best avoided by taking a closer look at the image. It is however possible that higher mammalian visual systems make some effort to deal with very high resolution information, and this article is offered mainly to show what kinds of things may be expected if they do.

The discussion falls into four parts. Firstly, there is the problem of finding the peaks in a mask response profile (Marr 1974b) in the case where they are close enough to interfere with one-another. Secondly, the result must be parsed into symbolic EDGE and BAR assertions. Thirdly, the detection of EDGE and BAR terminations is

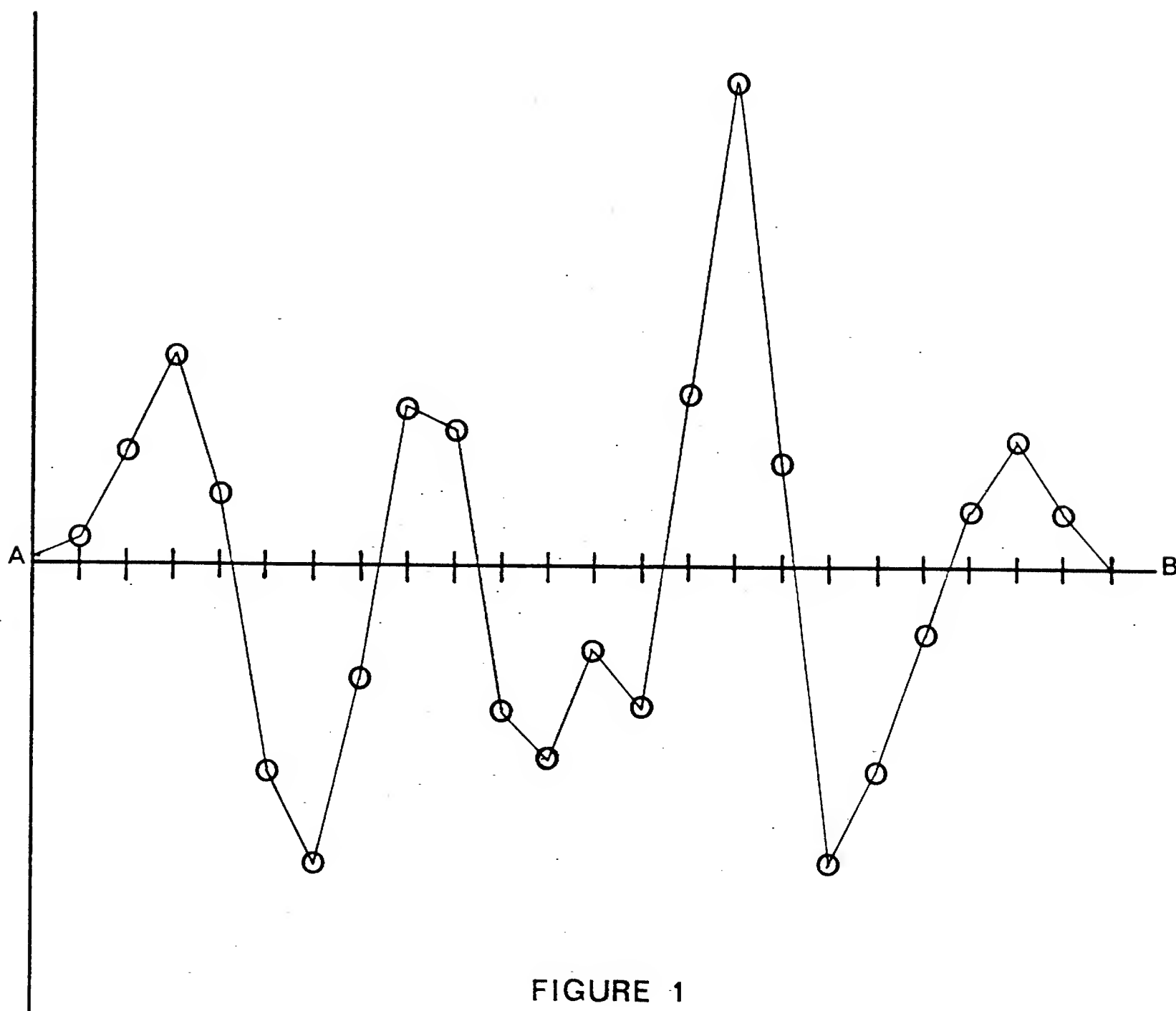
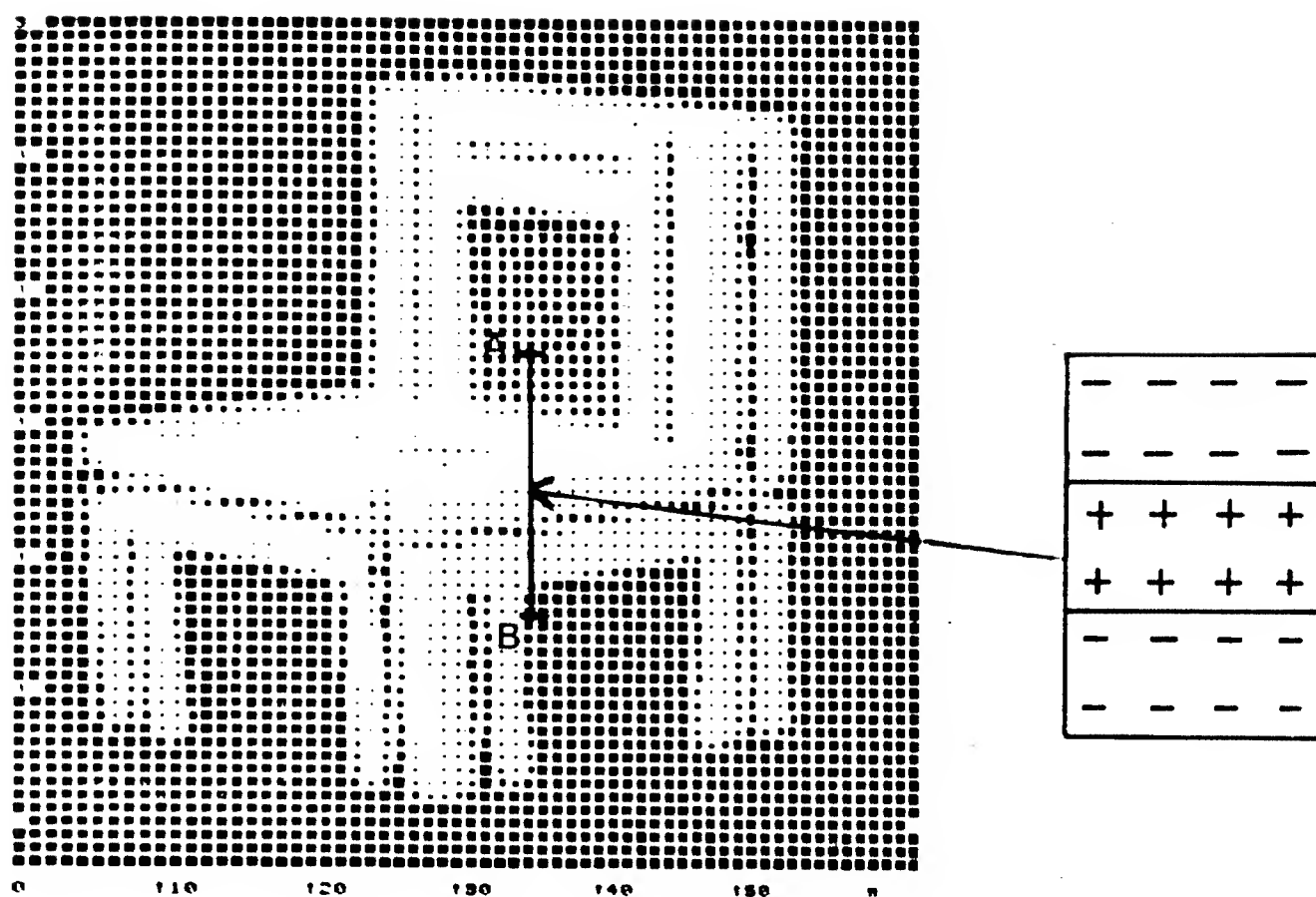
discussed; and fourthly, a method for detecting small points or blobs in the image is briefly mentioned.

Disentangling close peaks

After convolving a given mask with an image, the basic unit of data that one deals with is a sequence of numbers, representing the convolution at points equally spaced along a line perpendicular to the principal orientation associated with that mask. The sequence is terminated at either end by an expanse of unvarying intensity great enough so that parsing decisions within the sequence are independent of parsing decisions taken without it. An example of this appears in figure 1. The mask used there was bar-shaped, with a panel width of two image elements. It was thus exactly matched to the size of the "bars" in the image. The values shown in figure 1b were computed across quite a complex portion of the image, and were obtained at each image point. Interference due to the closeness of edges in the image is evident.

If one assumes that the edges in the image are sharp, then profiles like that in figure 1b may be regarded as being composed of a set of linearly smeared point sources. (The sharpness of the edges may be inferred from the size of and the difference between the values at neighbouring points.) These point sources may be recovered using linear decoding techniques in the following way. Let the measurements from a mask at points along a line perpendicular to the mask's orientation be $\{m(1), m(2), \dots, m(n)\}$, and suppose that all measurements made outside this part of the sequence are zero. The set of values in figure 1b is an

Figure 1. The small bar mask shown on the right, which has a panel-width of two image elements, was evaluated along the indicated path on this 64 by 64 intensity array. The result is shown in the graph below. The left-hand end of the graph corresponds to the bottom of the path of evaluation.



example of a set of $m(i)$. These values arise from a set of point sources which we may call $\{m^*(1), m^*(2), \dots, m^*(n)\}$, and the smearing relation between the $m(i)$ and the $m^*(i)$ is given by:

$$m(i) = 1/2m^*(i-1) + m^*(i) + 1/2m^*(i+1) \quad (1)$$

which holds at each point i . Isolating $m^*(i)$, the amount of point source at position i , we obtain:

$$m^*(i) = m(i) - 1/2(m^*(i+1) + m^*(i-1)) \quad (2)$$

Solving the family of simultaneous linear equations represented by (2) may be carried out by a matrix inversion, or by the parallel algorithm represented as a network in figure 2. If the distribution of weighting over the masks is not linear, but e.g. sinusoidal in structure, the only effect on the decoding network is to alter the coefficients from $(1/2, 1, 1/2)$ to whatever is appropriate: the technique is a general one. Note that this transform, which is reminiscent of (but not equal to) the inverse of the original measurement, is useful only because the image happens to be composed of sharp edges.

Linear interpolation between evaluation points

It will rarely be true that the edges in an image are conveniently positioned relative to the points at which the convolutions are obtained, and so it is important to ensure that an edge positioned between adjacent measuring points is represented in a sensible way (Marr 1974b). In the present case, intermediately positioned edges are represented by linear interpolation between adjacent measuring points, because de-smearing is a linear process; but the point is of sufficient

importance to warrant an explicit proof.

Interpolation lemma: Let the distance between measurements be 1, and suppose that there is a real edge of strength s in the image at a position d from one evaluation point P , and $(1-d)$ from its neighbour, Q . Then the output from the de-smearing process will place sources of strengths p at P , and q at Q , where $p + q = s$, and $d:(1-d) = p:q$.

Proof: Let the measurements made at P and at Q be m and n respectively. Because the process is linear, it suffices to analyse the transform in the case where the edge of strength s is the only item in the image. A mask of half-width 1, placed at a distance x away from the edge, will record a response of size $s(2-x)/2$. In particular, the measurement at P will be $s(2-d)/2$, and that at Q will be $s(1+d)/2$. The total contribution of the sources at P and at Q to the measurement at P is $(p+q/2)$: and at Q , it is $(q+p/2)$. Hence we see that

$$p + q/2 = s(2-d)/2, \text{ and}$$

$$p/2 + q = s(1+d)/2.$$

Hence $s = (p+q)$, $d = q/s$, and the lemma follows.

This result enables one to relate the distance between measurements directly to the resolution available from the results. On the assumption that between any two points at which measurements are taken there is only one source (i.e. edge) in the image, the position and strength of that source can be recovered exactly from measurements by masks of that size. Notice that if the decoding and parsing system were locally non-linear,

Figure 2. This network gives a parallel algorithm for solving the de-smearing problem (equations (2)). The conventions are as follows: unless otherwise stated, all connexions are linear. Open circles denote +, and filled circles, - inputs. The diamonds containing the fraction $1/2$ indicate that the quantity passing through it is halved. m and m^* are as defined in the text.

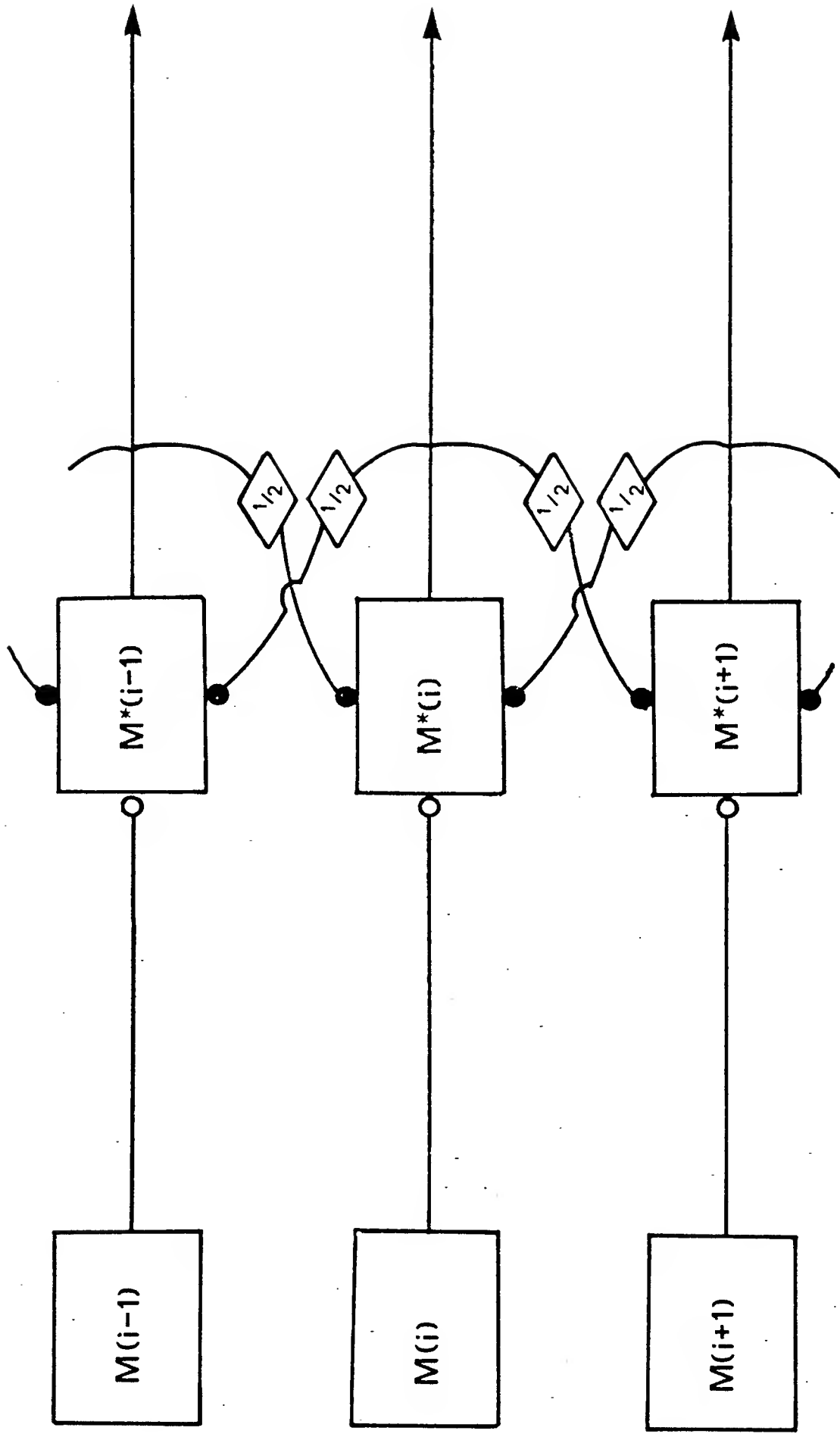


FIGURE 2

the apparatus for dealing with arbitrarily placed features in the image would have to be special.

The de-smearing process, and the reconstruction of point sources assuming linear interpolation, have been carried out on the profile of figure 1b, and the result appears in figure 4. One added piece of complexity was used to obtain the point sources A_i ; it affected only A_7 and A_8 , and is described next.

Parsing the de-smearred data

Once the list of point sources has been obtained, the question arises of how to parse them into a symbolic representation using the predicates BAR and EDGE. In the limiting situation that we are discussing, a BAR will be an edge-pair whose separation does not exceed the panel width of the smallest available mask; and other intensity changes will be described as EDGES. There are problems with this definition, because if there are more than two very close edges, one runs into what are essentially figure-ground problems in assigning the description (figure 3 has two parsings, for example), and the choice needs to be sensitive to a number of other factors. If one ignores this difficulty for the moment, and simply designs a method that will produce a sensible description of the image when such a description exists, or that is satisfied by all descriptions when more than one exists, one arrives at something like the following

METHOD: Let $A = \{A_i\}$ be the point sources obtained as described above. If the points sources are derived from "edge"-shaped masks, they correspond

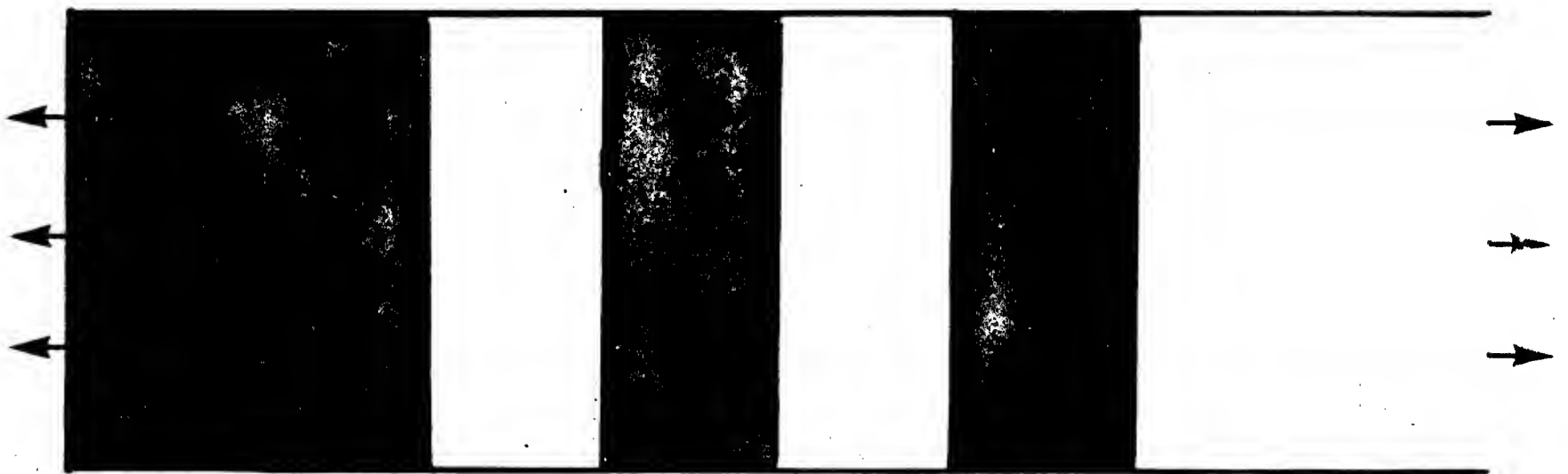


FIGURE 3

Figure 3. This image has two parsings: either there are two black bars to the right, and a black-to-white edge on their left; or there are two white bars to the left, and a black-to-white edge on their right. In this kind of situation, parsing decisions taken at one point have consequences that propagate across the image.

to edges in the image, and so EDGE assertions may be associated with each source in a one-to-one manner. If the point sources are derived from "bar"-shaped masks, they will be parsed into two patterns of sources, that corresponding to an edge in the image $(-x, +x)$, and that corresponding to a very thin line or bar $(-x, +2x, -x)$. The canonical method of parsing a bar-mask convolution is to use only the EDGE symbol, breaking the profile into $(-x, +x)$ patterns only. This avoids the figure-ground problem, and produces a unique output (provided that the underlying edges are sharp).

This method may be used to parse figure 4. If one uses the more complex parsing technique, the profile shown there is seen to contain, in order, an EDGE (arrows A1 & A2), a BAR (A3, A4, & A6), a shadow EDGE (A5 & A7), a dark BAR (A8, A9 & A10), and finally another EDGE (A11 & A12). This profile contains one further point of interest: A7 and A8 are so close that the value of the de-smearred output at the point lying between them is the sum of the contribution of each, and hence is very small. The neighbouring peaks cause the parsing algorithm to reconstruct A7 and A8 to be the correct size, because A9 and A10 force the existence of a BAR, which in turn requires A8 at the specified position. A7 is then defined because of the requirement that the point sources be compatible with the output from the de-smearing process.

Terminations

We have seen how to take account of the interactions between an edge or bar mask, and others of the same orientation that lie in a

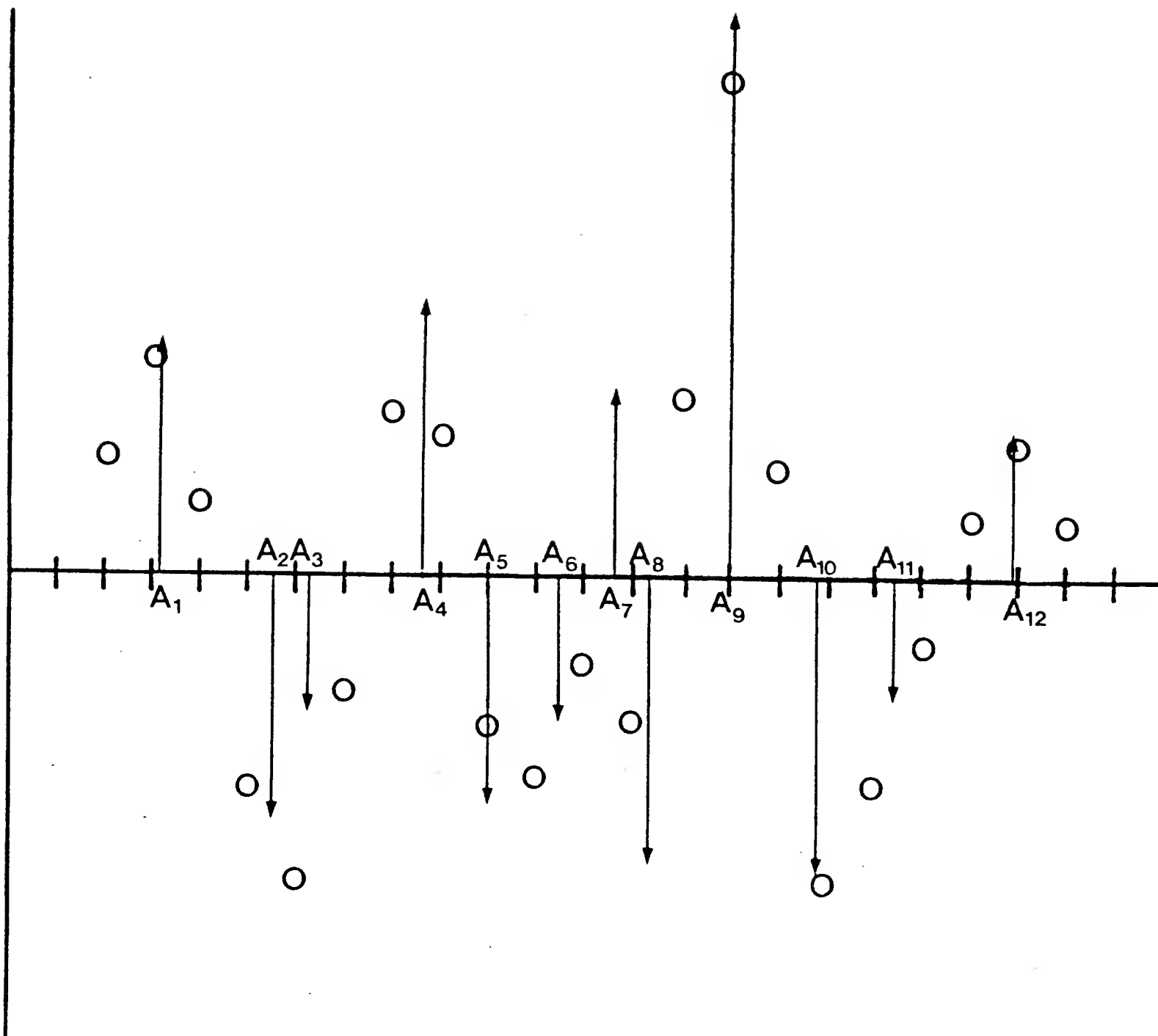


FIGURE 4

Figure 4. This figure gives the correct parsing of the sequence shown in figure 1. The A_i are the point sources, and their description is as follows: EDGE (A_1 , A_2); BAR (A_3 , A_4 , A_6); EDGE (A_5 , A_7); BAR (A_8 , A_9 , A_{10}); EDGE (A_{11} , A_{12}). The open circles show the original data.

direction perpendicular to that orientation. The interaction that ranks next in importance is longitudinal, between bars and edges of the same orientation that lie along that orientation.

The predicates EDGE and BAR may be regarded as signifying the amount of edge or of bar that is present in a region. We are interested in detecting terminations of edges and of bars, but to do this is not straightforward using EDGE and BAR assertions. The reason is that such assertions, as defined earlier, indicate the average amount of edge or bar over a considerable length, (the length of the mask in the original measurement), and a small gap in a bar would, for example, cause only a momentary dip in the strengths associated with the nearby BAR symbols: indeed, in the case of a strong bar separated from a weaker one by a small gap, it is very unclear from the distribution of the bar values that this, rather than a gradual fading away, is what is actually happening. It is however possible to analyse longitudinal interactions precisely, and in order to do this, it is convenient to regard an assertion, like those obtained earlier, as being composed of the sum along its length of elemental bar assertions which we shall denote by $b(j)$: figure 5a illustrates the idea. The new function b is in fact defined as follows:

$$\text{BAR}(i) = b(i-r) + b(i-r+1) + \dots + b(i+r-1) + b(i+r) \quad (3)$$

where $\text{BAR}(i)$ stands for the strength of the standard kind of BAR assertion made at the point i . The advantage of these new variables is that because they represent small pieces, terminations are easily characterised in terms of them. Notice that this kind of trick is only

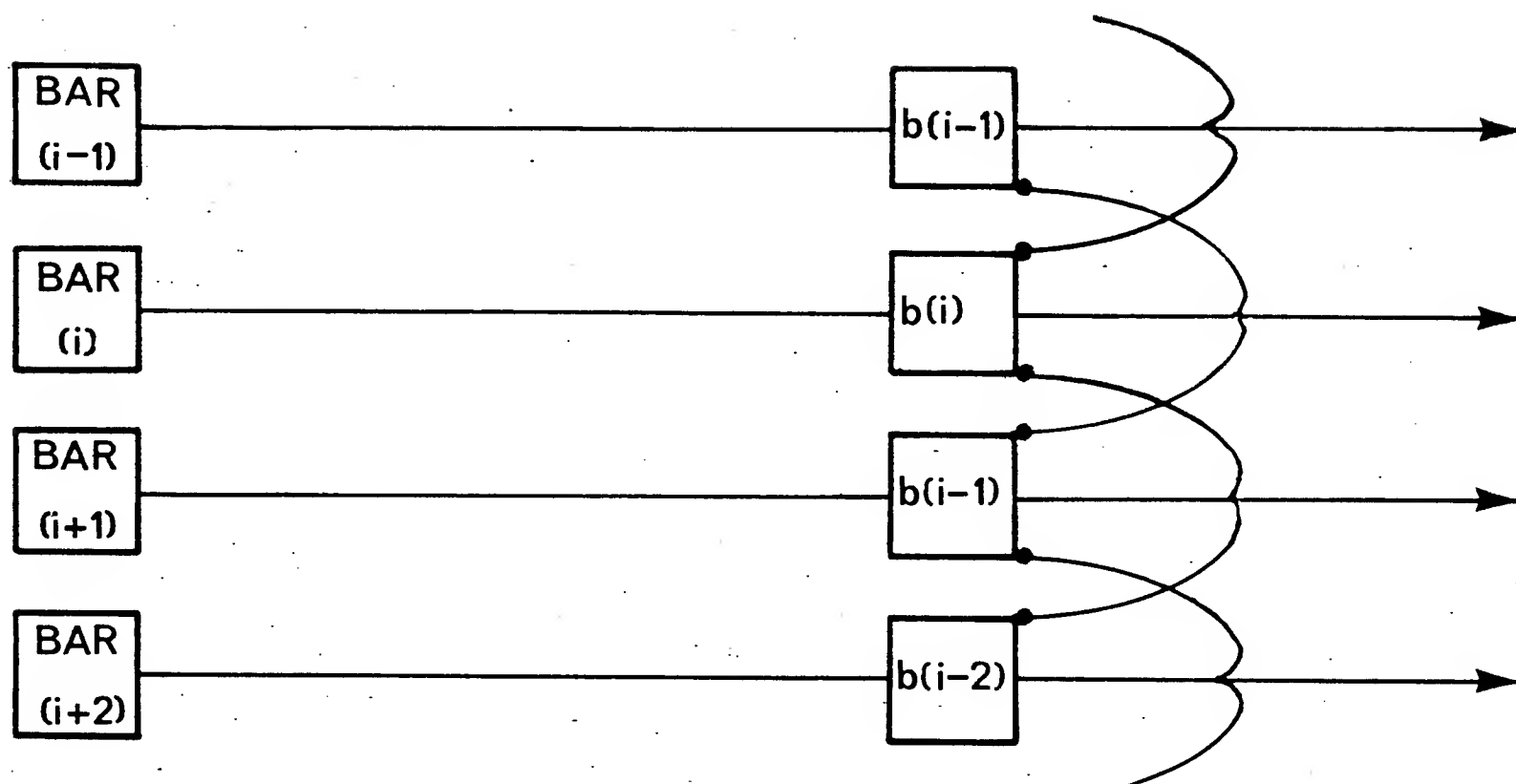
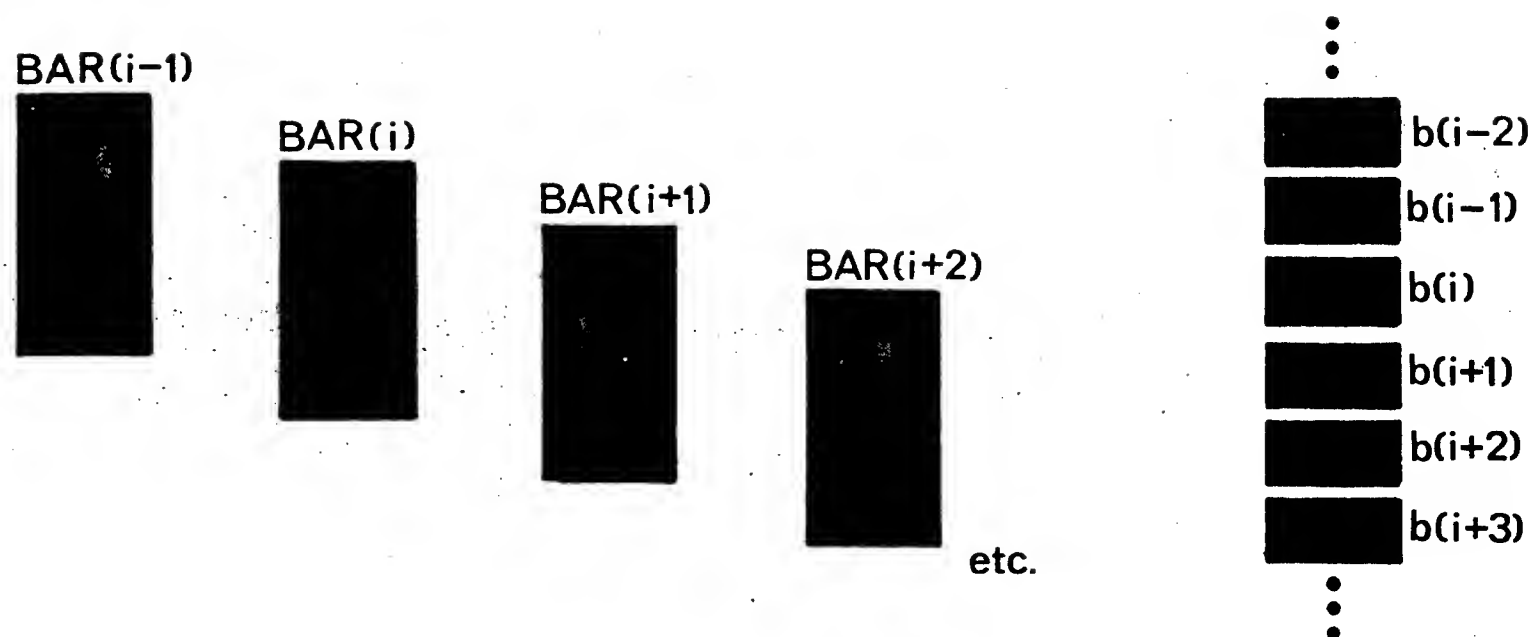


FIGURE 5

Figure 5. For the detection of terminations, each BAR assertion is imagined to be composed of 3 (or more) mini-assertions, $b(i)$. The network of 5b gives a parallel algorithm for recovering the $b(i)$ from the $\text{BAR}(i)$.

possible because we are dealing with assertions: it would have a complicated expression in terms of the original measurements. In principle, at a termination, one or more of the $b(i)$ is zero. (Gaps that are short, compared with the length of the $b(i)$, will give the same trouble with the elemental bars that larger gaps gave with the original bar assertions.) The length of the elemental bar units $b(i)$ therefore decides the size of gap that can be detected by this method, and this in turn is determined by the distance apart at which measurements should be made from which the bar assertions are computed. Extracting $b(i)$ from (3), we obtain:

$$b(i) = \text{bar}(i) - \sum_{j \neq 0} b(i+j) \quad (4)$$

These two independent sets of simultaneous equations are solved by a network like that of figure 5b. The assertion BAR-TERMINATION at this small scale may be defined by examining the values of the $b(i)$, and searching for places where b goes to zero. Similar techniques may be applied to EDGE assertions.

Recognising blobs

It was pointed out elsewhere (Marr 1974a) that methods for detecting blobs that rely upon masks with a centre-surround weighting distribution are either expensive or fallible. If the convolution from a bar-shaped mask is used, its signals will be difficult to interpret: but if BAR assertions are first computed, and terminations are derived from them, the result may be used rather easily to detect the presence of a very small blob. The criterion for the presence of such a blob is that a

short, doubly terminated BAR be present at all orientations at that point in the image. Notice that, as in the preceding case of the computation of terminations, this operation may easily be formulated in terms of BAR assertions, but it would be extremely clumsy if set out in terms of the original mask measurements.

The available evidence suggests however that we are unable to perceive small blobs that occupy less than about four receptors on the retina, (Cornsweet 1970 p356) and above this size, it becomes possible to talk in terms of the boundary of the blob. Further analysis of methods for detecting very small blobs is therefore almost certainly irrelevant, and I have not pursued the matter.

Acknowledgements: I thank Ken Forbus for programming assistance. Work reported herein was conducted at the Artificial Intelligence Laboratory, a Massachusetts Institute of Technology research program supported in part by the Advanced Research Projects Agency of the Department of Defense and monitored by the Office of Naval Research under Contract number N00014-70-A-0362-0005.

References

- Cornsweet, T.N. (1970). Visual perception. New York: Academic Press.
- Marr, D. (1974a) The purpose of low-level vision. MIT Artificial Intelligence Laboratory Memo 324.
- Marr, D. (1974b) The low-level symbolic representation of intensity changes in an image. MIT Artificial Intelligence Laboratory Memo 325.